

Minimal Size of Union

Yuanzhong Ou, Boli Wang, Min Yan
Hong Kong University of Science and Technology

August 15, 2011

1 Main Result

Given finite sets A_1, A_2, \dots, A_n with respective numbers a_1, a_2, \dots, a_n of elements, the union $A_1 \cup A_2 \cup \dots \cup A_n$ can have as many as $a_1 + a_2 + \dots + a_n$ elements and as few as $\max\{a_1, a_2, \dots, a_n\}$ elements. The maximum is realized when the sets are pairwise disjoint. When the minimum is realized, chances are there are many nonempty intersections among the sets.

In this paper, we fix $k \leq n$ and study the bound on the size of the union under the additional assumption that the intersection of any k sets is empty. For $k = 2$, this is the trivial pairwise disjoint case.

In a simpler version of the problem, the sets A_i are Lebesgue measurable subsets and the size of the subsets is the Lebesgue measure. The problem is simpler because any non-negative number is allowed to be the size, not just the non-negative integers.

Theorem 1. *Let $2 \leq k \leq n$ and let the non-negative numbers a_1, a_2, \dots, a_n be given. If A_1, A_2, \dots, A_n are Lebesgue measurable subsets, such that the intersection of any k subsets among A_i is empty and $\mu(A_i) = a_i$, then*

$$\mu(\cup A_i) \geq \max\{a_1, a_2, \dots, a_n\}, \quad \mu(\cup A_i) \geq \frac{1}{k-1}(a_1 + a_2 + \dots + a_n).$$

Moreover, it is always possible to find suitable A_i such that the bigger of the two lower bounds is realized.

The result remains true for any measure space (X, μ) with the property that $\mu(X) = \infty$, and for any $A \subset X$ of finite measure and any $b > 0$, there is a measurable $B \subset X$, such that $A \cap B = \emptyset$ and $\mu(B) = b$.

The counting version of the theorem is also true in general. We denote by $|A|$ the number of elements in A .

Theorem 2. *Let $2 \leq k \leq n$ and let the non-negative integers a_1, a_2, \dots, a_n be given. Let a be the smallest integer $\geq \frac{1}{k-1}(a_1 + a_2 + \dots + a_n)$. If A_1, A_2, \dots, A_n are finite sets, such that the intersection of any k sets among A_i is empty and $|A_i| = a_i$, then $|\cup A_i| \geq \max\{a_1, a_2, \dots, a_n, a\}$. Moreover, it is always possible to find suitable A_i , such that the lower bound $\max\{a_1, a_2, \dots, a_n, a\}$ is realized.*

2 The Lower Bound

The first lower bound is due to the fact that the union contains any A_i . Moreover, it is clear that $\mu(\cup A_i) = a_n$ if and only if A_n almost contains all the subsets.

For the second lower bound, we introduce “pure intersections” for distinct i_1, i_2, \dots, i_l

$$\begin{aligned} B_{i_1 i_2 \dots i_l} &= A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_l} - \cup_{j \neq i_1, i_2, \dots, i_l} A_j \\ &= A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_l} - \cup_{j \neq i_1, i_2, \dots, i_l} A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_l} \cap A_j. \end{aligned}$$

The theorem assumes $B_{i_1 i_2 \dots i_l} = \emptyset$ for $l \geq k$. Therefore we have disjoint union decompositions

$$\begin{aligned} A_j &= B_j \sqcup (\sqcup_{i \neq j} B_{ij}) \sqcup (\sqcup_{\substack{i_1, i_2 \neq j \\ i_1 < i_2}} B_{i_1 i_2 j}) \sqcup \dots \sqcup (\sqcup_{\substack{i_1, \dots, i_{k-2} \neq j \\ i_1 < \dots < i_{k-2}}} B_{i_1 \dots i_{k-2} j}), \\ A_1 \cup \dots \cup A_n &= (\sqcup_i B_i) \sqcup (\sqcup_{i_1 < i_2} B_{i_1 i_2}) \sqcup (\sqcup_{i_1 < i_2 < i_3} B_{i_1 i_2 i_3}) \sqcup \dots \sqcup (\sqcup_{i_1 < \dots < i_{k-1}} B_{i_1 \dots i_{k-1}}). \end{aligned}$$

Then

$$\begin{aligned} \mu(A_j) &= \mu(B_j) + \sum_{i \neq j} \mu(B_{ij}) + \sum_{\substack{i_1, i_2 \neq j \\ i_1 < i_2}} \mu(B_{i_1 i_2 j}) + \dots + \sum_{\substack{i_1, \dots, i_{k-2} \neq j \\ i_1 < \dots < i_{k-2}}} \mu(B_{i_1 \dots i_{k-2} j}), \\ \mu(A_1 \cup \dots \cup A_n) &= \sum_i \mu(B_i) + \sum_{i_1 < i_2} \mu(B_{i_1 i_2}) + \sum_{i_1 < i_2 < i_3} \mu(B_{i_1 i_2 i_3}) + \dots + \sum_{i_1 < \dots < i_{k-1}} \mu(B_{i_1 \dots i_{k-1}}). \end{aligned}$$

Adding the first equality together for various j and comparing with the second equality, we get

$$\begin{aligned} &\mu(A_1) + \mu(A_2) + \dots + \mu(A_n) \\ &= \sum_i \mu(B_i) + 2 \sum_{i_1 < i_2} \mu(B_{i_1 i_2}) + 3 \sum_{i_1 < i_2 < i_3} \mu(B_{i_1 i_2 i_3}) + \dots + (k-1) \sum_{i_1 < \dots < i_{k-1}} \mu(B_{i_1 \dots i_{k-1}}) \\ &\leq (k-1) \left(\sum_i \mu(B_i) + \sum_{i_1 < i_2} \mu(B_{i_1 i_2}) + \sum_{i_1 < i_2 < i_3} \mu(B_{i_1 i_2 i_3}) + \dots + \sum_{i_1 < \dots < i_{k-1}} \mu(B_{i_1 \dots i_{k-1}}) \right) \\ &= (k-1) \mu(A_1 \cup \dots \cup A_n). \end{aligned}$$

This gives the second lower bound. The proof also tells us that the second lower bound is realized if and only if

$$\mu(B_i) = \mu(B_{i_1 i_2}) = \mu(B_{i_1 i_2 i_3}) = \dots = \mu(B_{i_1 \dots i_{k-2}}) = 0.$$

This means that the pure intersections of j subsets are almost empty for any $j \neq k-1$. In other words, the elements of A_i are “concentrated” in the pure intersections of $k-1$ subsets.

3 Realizing the Lower Bound in Extreme Cases

We will use induction to construct the subsets that realize the lower bound. The construction for the case $k \leq n$ will use the realizability for the case $k-1 \leq n-1$ and the case $k \leq n-1$.

We can rely on $k - 1 \leq n - 1$ because the initial case $k = 2$ is trivial. If $k = 2$, then we always have $\mu(\cup_{i=1}^n A_i) = a_1 + a_2 + \cdots + a_n$, which is the lower bound.

The reliance on $k \leq n - 1$ is more complicated. The initial case is $k = n$. In fact, we will only need the case $k \leq n - 1$ when the second lower bound is bigger. So the initial case is covered by the following.

Proposition 3. *Suppose $a_i \geq 0$ satisfy*

$$\max\{a_1, a_2, \dots, a_n\} \leq \frac{1}{n-1}(a_1 + a_2 + \cdots + a_n). \quad (1)$$

Then there are Lebesgue measurable subsets A_i , such that

$$\cap_{i=1}^n A_i = \emptyset, \quad \mu(A_i) = a_i, \quad \mu(\cup_{i=1}^n A_i) = \frac{1}{n-1}(a_1 + a_2 + \cdots + a_n).$$

We expect the lower bound to be realized when all pure intersections are empty except the pure intersections of $n - 1$ subsets

$$C_i = A_1 \cdots (i-1)(i+1) \cdots n = A_1 \cap \cdots \cap A_{i-1} \cap A_{i+1} \cap \cdots \cap A_n.$$

So the construction is to find pairwise disjoint C_i and take

$$A_i = C_1 \sqcup \cdots \sqcup C_{i-1} \sqcup C_{i+1} \sqcup \cdots \sqcup C_n.$$

Let $x_i = \mu(C_i)$. Then we can find suitable C_i if and only if the system of linear equations

$$x_1 + \cdots + x_{i-1} + x_{i+1} + \cdots + x_n = a_i, \quad i = 1, 2, \dots, n,$$

has non-negative solution. The system has unique solution

$$x_i = \frac{1}{n-1}(a_1 + \cdots + a_n) - a_i, \quad i = 1, 2, \dots, n.$$

The condition for the solutions to be non-negative is exactly (1).

The argument gives the counting version of Proposition 3 when $\frac{1}{n-1}(a_1 + a_2 + \cdots + a_n)$ is an integer. In fact, the counting version of extreme case is true in general because of the following observation.

Remark. For realizing the lower bound in the counting version, in case the second lower bound is bigger, it is sufficient to show that the second lower bound can be realized when $a_1 + a_2 + \cdots + a_n$ is divisible by $k - 1$ (i.e., the second lower bound is an integer).

Here is the reason for the remark. Without loss of generality, we can always assume

$$a_1 \leq a_2 \leq \cdots \leq a_n. \quad (2)$$

If $a_1 = 0$, then the remark is really for the same k but with smaller n . The induction will reduce the remark to the initial case $k = n$. If we still have some $a_1 = 0$ in case $k = n$, then

$$a_n \geq \frac{1}{n-1}(a_2 + \cdots + a_n) = \frac{1}{n-1}(a_1 + a_2 + \cdots + a_n).$$

Therefore in the context of the remark, the two lower bounds must be equal and is an integer.

So we may further assume $a_1 > 0$ in addition to (2). Suppose $0 < r < k - 1$ is the remainder of the division of $a_1 + a_2 + \cdots + a_n$ by $k - 1$. Then

$$(a_1 - 1) + \cdots + (a_r - 1) + a_{r+1} + \cdots + a_n = a_1 + a_2 + \cdots + a_n - r$$

is divisible by $k - 1$, and

$$a_n \leq \frac{1}{k-1}(a_1 + a_2 + \cdots + a_n)$$

implies

$$a_n \leq \frac{1}{k-1}((a_1 - 1) + \cdots + (a_r - 1) + a_{r+1} + \cdots + a_n).$$

Suppose we have finite sets A'_1, \dots, A'_n , such that the intersection of any k sets is empty and

$$|A_i| = \begin{cases} a_i - 1, & \text{if } 1 \leq i \leq r, \\ a_i, & \text{if } r < i \leq n. \end{cases} \quad |\cup A'_i| = \frac{1}{k-1}((a_1 - 1) + \cdots + (a_r - 1) + a_{r+1} + \cdots + a_n).$$

Let $\langle 1 \rangle$ be a set of single element and take

$$A_i = \begin{cases} A'_i \sqcup \langle 1 \rangle, & \text{if } 1 \leq i \leq r, \\ A'_i, & \text{if } r < i \leq n. \end{cases}$$

Then the intersection of any k sets from A_i is empty, $|A_i| = a_i$, and

$$|\cup A_i| = |\cup A'_i| + 1 = \frac{1}{k-1}(a_1 + a_2 + \cdots + a_n - r) + 1$$

is the smallest integer $\geq \frac{1}{k-1}(a_1 + a_2 + \cdots + a_n)$.

4 Realizing the Lower Bound

We prove the realization part of Theorem 1. Without loss of generality, we assume (2).

We first consider the case the first lower bound is bigger. This means that

$$a_n \geq \frac{1}{k-2}(a_1 + a_2 + \cdots + a_{n-1}). \quad (3)$$

Note that the right side is the second lower bound for the case $k - 1 \leq n - 1$. As explained in Section 3, the induction can be used. We need to consider two subcases.

The first subcase is

$$a_{n-1} \geq \frac{1}{k-2}(a_1 + a_2 + \cdots + a_{n-1}).$$

By induction, we can find A_1, A_2, \dots, A_{n-1} , such that the intersection of any $k - 1$ subsets is empty, $\mu(A_i) = a_i$, and $\mu(\cup_{i=1}^{n-1} A_i) = a_{n-1}$. Let $\langle a \rangle$ be a subset of measure a and introduce

$$A_n = (A_1 \cup A_2 \cup \cdots \cup A_{n-1}) \sqcup \langle a_n - a_{n-1} \rangle.$$

Then among $A_1, A_2, \dots, A_{n-1}, A_n$, the intersection of any k subsets is empty, and $\mu(A_n) = \mu(\cup_{i=1}^n A_i) = \mu(\cup_{i=1}^{n-1} A_i) + (a_n - a_{n-1}) = a_n$.

The second subcase is

$$a_{n-1} \leq \frac{1}{k-2}(a_1 + a_2 + \dots + a_{n-1}).$$

By induction, we can find A_1, A_2, \dots, A_{n-1} , such that the intersection of any $k-1$ subsets is empty, $\mu(A_i) = a_i$, and $\mu(\cup_{i=1}^{n-1} A_i) = \frac{1}{k-2} \sum_{i=1}^{n-1} a_i$. The assumption (3) implies

$$\begin{aligned} b &= \frac{1}{k-1}(a_1 + a_2 + \dots + a_{n-1} + a_n) - \frac{1}{k-2}(a_1 + a_2 + \dots + a_{n-1}) \\ &= \frac{1}{k-1} \left(a_n - \frac{1}{k-2}(a_1 + a_2 + \dots + a_{n-1}) \right) \geq 0. \end{aligned} \quad (4)$$

So we may add

$$A_n = (A_1 \cup A_2 \cup \dots \cup A_{n-1}) \sqcup \langle b \rangle$$

to the list. Among $A_1, A_2, \dots, A_{n-1}, A_n$, the intersection of any k subsets is empty, and $\mu(A_n) = \mu(\cup_{i=1}^n A_i) = \mu(\cup_{i=1}^{n-1} A_i) + b = a_n$.

Now we turn to the case the second lower bound is bigger. This means that the number

$$b = \frac{1}{k-1}(a_1 + a_2 + \dots + a_n) - a_n \geq 0,$$

and we have

$$a_n = \frac{1}{k-1}((a_1 - b) + \dots + (a_{k-1} - b) + a_k + \dots + a_n).$$

If $b \leq a_1$, then for the problem of realizing the lower bound for n subsets of measure $a_1 - b, \dots, a_{k-1} - b, a_k, \dots, a_n$, such that the intersection of any k subsets is empty, the two lower bounds are equal. By the first case proved earlier, the first lower bound a_n (as well as the second one) can be realized by A'_1, A'_2, \dots, A'_n . In other words, we can find A'_i , such that the intersection of any k subsets is empty, and

$$\mu(A'_i) = \begin{cases} a_i - b, & \text{if } 1 \leq i < k, \\ a_i, & \text{if } k \leq i \leq n. \end{cases} \quad \mu(\cup A'_i) = a_n = \frac{1}{k-1}(a_1 + a_2 + \dots + a_n) - b.$$

Take

$$A_i = \begin{cases} A'_i \sqcup \langle b \rangle, & \text{if } 1 \leq i < k, \\ A'_i, & \text{if } k \leq i \leq n. \end{cases}$$

Then the intersection of any k subsets among A_i is still empty, $\mu(A_i) = a_i$, and

$$\mu(\cup A_i) = \mu(\cup A'_i) + b = \frac{1}{k-1}(a_1 + a_2 + \dots + a_n).$$

If $b \geq a_1$, then subtracting b from a_i may yield negative number. So we subtract a_1 instead and get

$$a_n \leq \frac{1}{k-1}((a_2 - a_1) + \dots + (a_{k-1} - a_1) + a_k + \dots + a_n).$$

This means that, in the problem of realizing the lower bound for $n - 1$ subsets of measure $a_2 - a_1, \dots, a_{k-1} - a_1, a_k, \dots, a_n$, such that the intersection of any k subsets is empty, the second lower bound is bigger. By Section 3, we may apply induction and find A'_2, A'_3, \dots, A'_n , such that the intersection of any k subsets is empty,

$$\mu(A'_i) = \begin{cases} a_i - a_1, & \text{if } 2 \leq i < k, \\ a_i, & \text{if } k \leq i \leq n, \end{cases}$$

and

$$\begin{aligned} \mu(\cup A'_i) &= \frac{1}{k-1}((a_2 - a_1) + \dots + (a_{k-1} - a_1) + a_k + \dots + a_n) \\ &= \frac{1}{k-1}(a_1 + a_2 + \dots + a_n) - a_1. \end{aligned}$$

Take

$$A_1 = \langle a_1 \rangle, \quad A_i = \begin{cases} A'_i \sqcup \langle a_1 \rangle, & \text{if } 2 \leq i < k, \\ A'_i, & \text{if } k \leq i \leq n. \end{cases}$$

Then the intersection of any k subsets from A_i is still empty, $\mu(A_i) = a_i$, and

$$\mu(\cup A_i) = \mu(\cup A'_i) + a_1 = \frac{1}{k-1}(a_1 + a_2 + \dots + a_n).$$

Finally, we explain how the proof can be adapted to Theorem 2.

Assume the first lower bound is bigger. In the first subcase, the construction is the same as before. In the second subcase, let a' be the smallest integer $\geq \frac{1}{k-2}(a_1 + a_2 + \dots + a_{n-1})$. Then the assumption (3) implies the inequality (4), so that $a \geq a'$. By induction, we have A_1, A_2, \dots, A_{n-1} , such that the intersection of any $k - 1$ subsets is empty, $|A_i| = a_i$, and $|\cup_{i=1}^{n-1} A_i| = a'$. Adding $A_n = (A_1 \cup A_2 \cup \dots \cup A_{n-1}) \sqcup \langle a - a' \rangle$ to the list completes the construction.

Next assume the second lower bound is bigger. As remarked in Section 3, we may further assume that $\frac{1}{k-1}(a_1 + a_2 + \dots + a_n)$ is an integer. Then all the numbers appearing in the rest of the construction are integers. The proof is complete.